DOCUMENT RESUME

ED 091 423 TM 003 637

AUTHOR	Feir, Betty J.; Toothaker, Larry E.
TITLE	The ANOVA F-Test Versus The Kruskal-Wallis Test: A Robustness Study.
PUB DATE	[Apr 74]
NOTE	38p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Analysis of Variance; Comparative Statistics; *Nonparametric Statistics; Sampling; Statistical Analysis; Statistical Studies
IDENTIFIERS	ANOVA F Test; Kruskal Wallis Test; *Variance (Statistical)

ABSTRACT

Researchers are often in a dilemma as to whether parametric or nonparametric procedures should be cited when assumptions of the parametric methods are thought to be violated. Therefore, the Kruskal-Wallis test and the ANOVA F-test were empirically compared in terms of probability of a Type I error and power under various patterns of mean differences in combination with patterns of variance inequality, and patterns of sample size inequality. The Kruskal-Wallis test was found to be competitive with the ANOVA F-test in terms of alpha but not for power. Power of the Kruskal-Wallis test was grossly affected in all but one situation for nonstepwise mean differences when sample sizes and variances were negatively related and when small levels of significance were utilized. The ANOVA F-test, however, was found to be generally robust for the types of specified mean differences. (Author)

Session # 21.8.

THE ANOVA F-TEST VERSUS THE KRUSKAL-WALLIS TEST: A ROBUSTNESS STUDY

ABSTRACT

Betty J. Feir, Oklahoma State University

Larry E. Toothaker, University of Oklahoma

Researchers are often in a dilemma as to whether parametric or nonparametric procedures should be cited when assumptions of the parametric methods are thought to be violated. Therefore, the Kruskal-Wallis test and the ANOVA F-test were empirically compared in terms of probability of a Type I error and power under various patterns of mean differences in combination with patterns of variance inequality, and patterns of sample size inequality. The Kruskal-Wallis test was found to be competitive with the ANOVA F-test in terms of alpha but not for power. Power of the Kruskal-Wallis test was grossly affected in all but one situation for nonstepwise mean differences when sample sizes and variances were negatively related and when small levels of significance were utilized. The ANOVA F-test, however, was found to be generally robust for the types of specified mean differences.

THE ANOVAF-TEST VERSUS THE KRUSKAL-WALLIS TEST:

A ROBUSTNESS STUDY

Betty J. Feir, Oklahoma State University
Larry E. Toothaker, University of Oklahoma

A common problem in applied research is to decide whether or not sample differences in central tendency reflect true differences in parent populations. It is appropriate to use the one-way fixed effects ANOVA F-test for the K-sample case ($K \geq 2$) if assumptions of normality, homogeneity of variances, and independence of errors are met. When either normality or equality of variances is doubtful, the use of nonparametric statistical procedures is often recommended (Bradley, 1968).

The two-sample Mann-Whitney U test is a frequently used ditribution-free analogue to the Student's t-test, and in a one-way ANOVA situation, the distribution-free analogue to the F-test is the Kruskal-Wallis rank test (Kruskal, 1952). Both of these distribution-free methods have been commonly used to test hypotheses about means. In fact, they are generally sensitive to differences in location, but specifically sensitive to differences in medians. If the popluations are symmetric, then the means and the medians are the same. For location differences, the Kruskal-Wallis test as compared to the F-test has an symptotic relative efficiency of .95 for the normal case, 1.00 for the uniform case, and a lower bound of .864 (Bradley, 1968). Thus, not more than 13.6% asymptotic efficiency can be lost using the Kruskal-Wallis test rather than the F-test, and, if the distribution is normal, the asymptotic loss of efficiency is only about 5%. Keeping in mind that asymptotic relative efficiencies are computed for unrealistically large sample sizes with minusclue differences in measures of location, it would seem profitable to investigate the efficiency of the Kruskal-Wallis test as compared to the F-test for realistic sample sizes and realistic differences in location (Feir and Toothaker, 1974).

Research has demonstrated, however, that in certain cases the ANOVA F-test as well as other parametric procedures can be said to be "robust" or insensitive to assumption violation (Box and Andersen, 1955). Robustness creates a dilemma for the researcher: should he or should he not utilize nonparametric procedures in lieu of parametric procedures? Many studies have investigated the effect of assumption violation on parametric procedures, but very few have directly compared parametric to nonparametric techniques under assumption violation. Nevertheless, literature often recommends distribution-free techniques as being powerful alternatives when parametric assumptions are violated (Terry, 1952; Hodges and Lehman, 1961; Klotz, 1963; Van der Laan, 1964; Puri, 1964; McSweeney, 1967; Bradley, 1968; Penfield and McSweeney, 1968; McSweeney and Penfield, 1969). Since under assumption violation the F-test appears to be robust and the Kruskal-Wallis test appears to be powerful, perhaps the two tests should be compared under extensive combinations of parametric assumption violations. The present research empirically investigates the effect of non-normality coupled with a variety of unequal variance conditions, and several types of mean differences for the F-test and the Kruskal-Wallis test for five levels of significance. Sample size per treatment level as well as total sample size are manipulated along with the assumption violations.

PROCEDURE

The basic numbers upon which this Monte Carlo study was based were obtained via a computer using a pseudo-random number generator. Depending upon the assumption violation, the numbers were selected from either a normal distribution or from an exponential distribution scaled to have equal medians of zero value under the null hypothesis. The random deviates were allocated to four treatment levels which comprised a one-way fixed effects analysis of variance situation.

The observations from the normal distribution were derived by a technique developed by Box and Muller (1958), which generates pseudo-random variables distributed N(0,1). For the null situation, the means of the four treatment levels were zero. The non-null situation was established by defining values of μ_j (see Table 1), j = 1,2,3,4; such that the power for the ANOVA F-test would be about .86 for large mean differences and about .60 for smaller mean differences for the equal variances condition for the normal distribution. For each of the theoretical power values, both stepwise (1:2:3:4) and nonstepwise (1:4:4:4) differences were determined. Specification of the μ_j's for the normal distribution was made through the use of the non-centrality parameter, θ, (Pearson and Hartley, 1951) where,

$$\theta = \sqrt{\frac{\Sigma n_j \quad (\mu_j - \mu)^2}{J\sigma_e^2}}$$

Setting $\sigma_e^2 = 1$ and $J = 4$, and using probability of a Type I error equal to .05, the values of μ_j were found in a stepwise and a nonstepwise manner for both power values. Since equal sample size cases and unequal sample size cases would lead to different values of μ_j, the values were calculated for both equal and unequal sample sizes.

- -

Insert Table 1 about here

- -

The mean differences were used for the non-null situation for all variance conditions.

Four types of variances were established as follows: 1) equal variances (1:1:1:1); 2) unequal variances, stepwise (1:2:3:4); 3) unequal variances, with one small variance and three large variances, (SLLL, in the ratio of 1:4:4:4); 4) unequal variances with three small variances and one large variance (SSSL, in the ratio of 1:1:1:4). The average variance in all four cases was approximately equal to unity.

- -

Insert Table 2 about here

- -

The exponential distribution was derived by a method given by Lehman and Bailey (1968):

for $f(t) = pe^{-pt}$, with $p = 1$,

$E(t) = 1/p = 1$, and $\mathrm{var}(t) = 1/p^2 = 1$. Pseudo-random exponential variables were generated by multiplying the negative of the mean, $-E(t) = -1$, times the natural logarithm of uniform random variates distributed on the unit interval (IBM, 1968). The exponential variates were then scaled so that the medians would be zero. The resulting skewed population had median equal to zero, a variance of σ_j^2, a skewness measure of $\gamma_1 = 2$, and a kurtosis measure of $\gamma_2 = 6$. For this distribution the medians have zero value under the null hypothesis, but the values of means are non-zero. For equal variances, the value of the means was .30685. For unequal variances, the mean for any group j is .30685 σ_j. In order to simulate null and non-null conditions in the exponential distribution the values of the medians for group j were identical to the values of μ_j used in the normal distribution as shown in Table 1. The variances were also identical to those used in the normal distribution as shown in Table 2.

Comparisons between the ANOVA F-test and the Kruskal-Wallis test for equal sample sizes and a total sample size of 28 were made on all four variance cases for both the normal distribution and the exponential distribution. For three types of unequal sample sizes {(4,6,8,10), (4,8,8,8), (6,6,6,10)}, total sample size of 28, comparisons were made on all four variance cases which were both positively and negatively related to sample sizes for both the normal distribution and the exponential distribution. In addition, for a total sample size of 68, two types of unequal sample sizes {(11,15,19,23), (11,19,19,19)} were used for comparisons of the F-test and the Kruskal-Wallis with three types of variance inequality which were both positively and negatively related to sample sizes for both distributions. For each comparison of the F-test to the Kruskal-Wallis Test, 1,000 experiments were performed, where an experiment consisted of computation of each statistical test. The proportion of rejections in 1,000 experiments when there were no location differences was determined by comparing the empirical value of each test to five theoretical levels of significance, (.10, .05, .025, .01, .005) and each was referred to as empirical alpha. For each empirical alpha, four cases of empirical power were calculated by observing the proportion of rejections when each of the four types of differences in location was specified. (Table 15 summarizes the conditions which were investigated.)

It should be noted that the equal sample size, equal variance case for the normal distribution was included in the present study for the purpose of establishing validity of the Monte Carlo method.

RESULTS

Equal Sample Sizes {7,7,7,7}

Normal population (Table 3): When all assumptions are met, both the ANOVA F-test and the Kruskal-Wallis test provide good approximations to theoretical alpha. However, F-test approximations are somewhat closer to theoretical alpha than those of the Kruskal-Wallis test in the majority of cases. As is expected in this situation, the power of the F-test is higher than that of its nonparametric counterpart for all variance cases. In addition, the F-test displays only slight variability between power of stepwise mean differences and power of nonstepwise mean differences. The Kruskal-Wallis test, on the other hand, is greatly affected by the type of mean differences and usually displays a predominately higher power for stepwise differences than for nonstepwise differences. An illustration comparing power for the F-test and the Kruskal-Wallis test for a normal distribution with equal sample sizes for stepwise inequality of variances is shown in Figure 1, which is based on data from Table 3.

Exponential Population (Table 4): When sampling from an exponential population with equal sample sizes, the empirical alphas of both tests for unequal variances cases more closely approximate theoretical alpha for the smaller values of alpha than when variances are equal. For all cases of variance inequality, empirical alphas of the ANOVA F-test and the Kruskal-Wallis test fell within two standard deviations of a proportion to theoretical alpha. Therefore, in terms of alpha alone, neither test could be recommended over the other.

With regard to power, discrimination between the two statistical procedures is evident. The Kruskal-Wallis test is generally a more powerful procedure than the ANOVA F-test for this case. When differences in means are stepwise, for every variance case and at all significance levels but one, the Kruskal-Wallis test is slightly more powerful than the F-test. Nonetheless, an interesting pattern of power values erupts for nonstepwise mean differences. In the extreme tail regions (α=.01, α=.005), where Bradley (1968) indicates that parametric techniques would not be powerful, the ANOVA F-test is more powerful than the Kruskal-Wallis test in every instance of nonstepwise differences in means. This again illustrates the differential sensitivity (especially in remote tail regions) of the Kruskal-Wallis test to the type of location differences specified. It should be noted (see Figure 2 based on Table 4) that as the level of significance gets smaller, the variability between the power of stepwise and nonstepwise mean differences becomes smaller for the ANOVA F-test and larger for the Kruskal-Wallis test.

<u>Unequal Sample Sizes</u> {4,6,8,10}

Normal population (Table 5): When variances are equal, there is little difference between the empirical alphas of the ANOVA F-test and those of the Kruskal-Wallis test. For all levels of significance, except α=.10, empirical values for equal variances are within sampling variability for both tests. With positively related sample sizes and variances, both the ANOVA F-test and the Kruskal-Wallis test are conservative (as expected from normal theory), with the ANOVA F-test exemplifying much closer approximations to all five theoretical alphas than does the Kruskal-Wallis test. There appears to be little difference in the empirical alpha values of either test as to the type of variance inequality for positive relationships with sample sizes (for ≠ variances, 1:4:4:4, the alphas are slightly larger than for 1:1:1:4 and 1:2:3:4). However, for negatively related sample sizes and variances, where normal theory expectations are liberal, the Kruskal-Wallis test exhibits conservativism for both stepwise unequal variances and for unequal variances in the ratio of 1:4:4:4 in the extreme tail regions. Despite the conservativism in some instances, the empirical alpha values for the Kruskal-Wallis test more closely approximate theoretical alphas than do those of the F-test for all negatively related cases. In addition, the F-test makes more Type I errors than does the Kruskal-Wallis test for all negative relationships between sample sizes and variances. Solely on the basis of alpha approximations, the Kruskal-Wallis test appears to be the better technique for negative relationships between sample sizes and variances, but a look at the empirical power values changes the perspective.

The power of both tests for nonstepwise mean differences is consistently less than the power for stepwise differences when sample sizes and variances are negatively related (See Figure 3, based on Table 5). Further, the power of the Kruskal-Wallis test is always lower than that of the ANOVA F-test for both types of mean differences, and reduces nearly to zero for nonstepwise differences established for small theoretical alphas (when α=.005; $(1-\beta)_{NSL}$ = .047, .026, .042). In the positively related sample sizes and variances cases where alpha approximations for both tests are good, the power of the Kruskal-Wallis test is still considerably less than the power of the ANOVA F-test. For a normal distribution with stepwise unequal sample sizes, if inequality of variances is suspected, the researcher should recognize the deficiency in power of the Kruskal-Wallis test.

Exponential Population (Table 6): Apparently, the addition of non-normality to unequal sample sizes of {4,6,8,10} has little effect on the empirical alphas of the ANOVA F-test and the Kruskal-Wallis test. These results from the exponential population are similar to the preceding section for a normal distribution in that the F-test provides generally closer approximations to theoretical alpha for positively related cases of sample size and variance inequality and the Kruskal-Wallis test provides closer approximations for negatively related cases. For stepwise inequality of variances positively related to sample sizes, the empirical alphas of both tests are generally closer to theoretical alphas than those of the normal distribution, and for negative relationships they are slightly farther away from theoretical alphas than those of the normal distribution. The same unexpected conservativism occurs for the Kruskal-Wallis test when sample sizes and variances are negatively related for only two cases of theoretical alpha.

Sampling from an exponential population does have a different effect on power. For positively related sample sizes and variances, the power of the Kruskal-Wallis test is comparable to or larger than the power of the ANOVA F-test in all cases of unequal variances for both types of mean differences. Neither the Kruskal-Wallis test nor the ANOVA F-test shows any particular trend in power between stepwise and nonstepwise mean differences for the positively related cases. But with negatively related cases of sample sizes and variances, the power of the Kruskal-Wallis test once again behaves strangely (See Figure 4 based on Table 6). When nonstepwise mean differences are specified, they are, as shown in previous results, a deterrent to the use of the Kruskal-Wallis test in negatively related cases for small theoretical alphas.

As seen in Table 6, the power of the Kruskal-Wallis test approaches zero for nonstepwise differences in all variance cases. Throughout the entire range of theoretical alphas, stepwise mean differences generate larger power values for both tests than do nonstepwise mean differences in the negatively related sample sizes and variances cases. For small theoretical alphas, the Kruskal-Wallis test, even for stepwise mean differences, produces power values less than those of the ANOVA F-test in a few instances for the negatively related cases. (See Table6; $\neq$ n's, $\neq \sigma_e^2$; SSSL; α = .01, .005). The power of the ANOVA F-test for both stepwise and nonstepwise differences does remain fairly stable for all levels of theoretical alpha. Due to the instability and lack of power for the Kruskal-Wallis test with negatively related sample sizes and variances, the researcher should exercise extreme caution in utilizing this nonparametric procedure.

<u>Unequal Sample Sizes</u> {4,8,8,8} {6,6,6,10}

Due to the multiplicity of data, and the similarity of results, these two cases of *unequal* sample sizes will be grouped together in a generalized commentary.

Normal Population (Tables 7 and 9): When variances are equal, approximations to theoretical alpha for both tests and both sample sizes are generally good. For the positively related cases of sample sizes and variance inequality, both tests are conservative in most cases. The ANOVA F-test and the Kruskal-Wallis test do show an unexpected distinct reversal from normal theory in a few cases {(Table 7; $\neq$ n's, $\neq \sigma_e^2$; SSSL) (Table 9; $\neq$ n's, $\neq \sigma_e^2$; SLLL)}, but all are within sampling variability of theoretical alpha. For the negatively related cases, the Kruskal-Wallis test is usually conservative when a liberal test is expected {(Table 7; $\neq$ n's, $\neq \sigma_e^2$; Stepwise, SSSL, SLLL) (Table 9; $\neq$ n's, $\neq \sigma_e^2$;

Stepwise, SSSL, SLLL)}. There are scattered instances where the ANOVA F-test is also unexpectedly conservative, but in all cases of this type the ANOVA F-test is less conservative than is the Kruskal-Wallis test. These reversals from normal theory could very well be a function of the type of sample size inequality in conjunction with the type of variance inequality. Unanticipated reversals of this type create a problem for the researcher and further investigation in this area is needed. At best, if the Kruskal-Wallis test is used, the researcher with sample sizes of this type should expect a conservative test no matter what type of variance inequality is suspected. In using the ANOVA F-test, normal theory holds in the majority of cases.

When variances are equal and when variances are positively related to sample sizes, there is almost no difference between power of stepwise mean differences and nonstepwise mean differences for either test. Generally, for these variance cases, power of the ANOVA F-test is comparable to or better than that of the Kruskal-Wallis test. Certainly there is no indication that nonparametric power is significantly more enticing than parametric power or _vice versa_.

In the negatively related sample sizes and variances cases, there is, as shown throughout the results, an obvious difference in power between the two statistical procedures. Stepwise mean differences produce higher power than nonstepwise mean differences for both tests, with the ANOVA F-test having better power in practically all cases. Empirical power values of the Kruskal-Wallis test for nonstepwise mean differences fall to almost zero for small set alphas, and large nonstepwise differences reduce to almost the same empirical power values as those produced by saml1 nonstepwise mean differences. In other words, whether large or small nonstepwise mean differences are specified, the power is comparable. Apparently, large nonstepwise mean differences for the Kruskal-

Wallis test are severely affected in terms of power for small alphas when variances and sample sizes are negatively related (Seed Figure 5 based on Table 7). Overall, the ANOVA F-test seems to be a better technique in terms of power.

Exponential Population (Tables 8 and 10): When variances are equal the Kruskal-Wallis test is slightly conservative for all theoretical alpha values, with one expection, for both sample sizes. The ANOVA F-test, however, provides empirical alphas that are very close to theoretical alphas for sample sizes of {4,8,8,8} and are slightly conservative for sample sizes of {6,6,6,10}. For small values of theoretical alpha, the ANOVA F-test again shows a reversal from normal theory when variances and sample sizes are positively related. {(Table 8; ≠ n's, ≠ σ_e^2; Stepwise, SLLL) (Table 10; ≠ n's, ≠ σ_e^2; SSSL)}. The Kruskal-Wallis test is conservative in all cases of positively related sample sizes and variances. Empirical alphas for the negatively related cases for both of these sample sizes are considerably closer to theoretical alpha than were those of the stepwise unequal sample size cases ({4,6,8,10}; Table 6). In fact, empirical alphas of both tests are consistently within two standard deviations of a proportion difference of theoretical alpha.

As far as power is concerned, the Kruskal-Wallis test displays the same lack of power for nonstepwise mean differences for small values of theoretical alpha that it has displayed in previous cases. (See Figure 6 based on Table 8). For stepwise mean differences, there is only a small difference between the power of the ANOVA F-test and the power of the Kruskal-Wallis test, with the power of the Kruskal-Wallis test being slightly higher in most cases. Nevertheless, the undependability of empirical power of the Kruskal-Wallis test in certain situations prevents a recommendation for use as a more powerful technique than the ANOVA F-test.

<u>Unequal Sample Sizes</u> {11,15,19,23} {11,19,19,19}

Normal Population (Tables 11 and 13): For positively related sample sizes and stepwise unequal variances, empirical alphas of the ANOVA F-test and the Kruskal-Wallis test are very similar. The empirical alphas for sample sizes of {11,19,19,19} are slightly closer approximations to set alpha than are empirical alphas for sample sizes of {11,15,19,23}, but the differences are small. When sample sizes and variances are negatively related, the stepwise increment in sample sizes of {11,15,19,23} provides closer approximations to set alphas than does the nonstepwise increment {11,19,19,19}. Although both the Kruskal-Wallis test and the ANOVA F-test are both liberal for both sample sizes in the negative related cases, sample sizes of {11,15,19,23} produce much more liberal tests than do sample sizes {11,19,19,19} with empirical alphas of the Kruskal-Wallis test most closely approximating theoretical alphas in most instances. In fact, throughout the majority of cases investigated for these sample sizes with a normal distribution and unequal variances, the empirical alphas of the Kruskal-Wallis test more closely approximate theoretical alphas than do the empirical alphas of the ANOVA F-test. It should be noted that both tests show departures from normal theory expectation in specific instances (Table 13; $\neq$ n's, $\neq \sigma_e^2$; positively related, SSSL; and Table 11; $\neq$ n's, $\neq \sigma_e^2$; negatively related, 1:2:3:4).

In terms of power, stepwise mean differences generally produce higher empirical power than nonstepwise mean differences for both statistical tests using both types of sample size inequality. The power of the Kruskal-Wallis test in the majority of cases is lower than that of the ANOVA F-test. As demonstrated throughout this study, for negatively related cases, nonstepwise mean differences, the power of the Kruskal-Wallis test is especially low, particularly in remote tail regions. Even though the Kruskal-Wallis test would be the recommneded procedure in terms of alpha, the power conscious researcher testing at small levels of significance would be seriously handicapped utilizing the nonparametric procedure.

Exponential Population (Tables 12 and 14): There is surprisingly little difference between the empirical alphas of the Kruskal-Wallis test and the ANOVA F-test for either sample size investigated. Furthermore, there appears to be only a negligible effect due to the type of sample size inequality for the large total sample size of 68 coupled with nonnormality. Both tests show departive from normal theory conservativism at all alpha levels when variances are in the ratio of 1:1:1:4 for positively related sample size and variance inequality (Tables 12 and 13). This was the only variance condition which showed deviation from normal theory expectation. In general, for negatively related cases of sample size and variance inequality, the Kruskal-Wallis test provides slightly better approximations to theoretical alpha than does the ANOVA F-test but the differences are small for all but one type of variance inequality (SSSL). In this case, the ANOVA F-test appears to make considerably more Type I errors than does the Kruskal-Wallis test. In terms of Type I errors, the Kruskal-Wallis test might be the preferred procedure, but the differences in alpha approximations are not generally large.

In terms of power, stepwise mean differences produce higher empirical power than nonstepwise mean differences for both statistical tests using both types of sample size inequality. There is a smaller differential between empirical power based on stepwise differences and that based on nonstepwise differences for these larger sample sizes coupled with nonnormality than for the smaller sample sizes previously discussed. The Kruskal-Wallis test, even for nonstepwise mean differences with negatively related sample sizes and variances, generally demonstrates slightly higher power than that of the ANOVA F-test. However, the differences between power of the two statistical procedures is not large in any case, and one test could not be recommended over the other in terms of power.

SUMMARY AND CONCLUSIONS

The Kruskal-Wallis test has often been recommended for use in lieu of the ANOVA F-test when assumptions of the parametric procedure are violated. On the other hand, research has shown the ANOVA F-test to be relatively robust to a variety of assumption violations. To better guide the researcher, the present study empirically compared the robustness of the Kruskal-Wallis test and the ANOVA F-test in terms of probability of a Type I error and power under an assortment of varying conditions.

Upon comparing empirical alphas of the two statistical procedures, it was found for equal variances and for positively related sample sizes and variances that the ANOVA F-test provided closer approximations to theoretical alphas in the majority of cases for both normal and exponential distributions for all combinations of sample size inequality when N=28. When sample sizes were equal, the two procedures were comparable with the Kruskal-Wallis test providing slightly better alpha approximations for the exponential distribution and the ANOVA F-test providing slightly better alpha approximations for the normal distribution. When sample sizes and variances were negatively related, the empirical alphas of the Kruskal-Wallis test generally more closely approximated theoretical alphas for both distributions and all combinations of sample size inequality but one ({10,6,6,6}, Normal Population). For larger unequal sample sizes (N=68), the Kruskal-Wallis test generally demonstrated slightly closer approximations to theoretical alpha for the normal distribution and the exponential distribution.

For both statistical tests when sample sizes and variances were unequal, reversals from normal theory alphas were noted in some cases. Both tests were conservative for negative relationships and liberal for positive relationshipd in a few cases. These reversals were not found to be particularly illustrative of a trend toward poor estimates of theoretical alpha but rather simply indicative of specific instances of deviation from normal theory expectancy and warrant mention.

On the basis of alpha approximations alone, the Kruskal-Wallis test was found generally to be competitive with the ANOVA F-test when parametric assumptions were violated, but the Kruskal-Wallis test could not be recommended as the preferred procedure.

Power comparisons indicated that when sampling from a normal distribution with both positively and negatively related sample sizes and variances, the empirical power of the ANOVA F-test was higher than that of the Kruskal-Wallis test. When positive relationships of sample sizes and variances were present for the exponential distribution,the Kruskal-Wallis test generally demonstrated slightly higher power than that of the F-test. However, for negative relationships, with sampling from an exponential distribution the Kruskal-Wallis test consistently displayed higher power than the F-test only when stepwise mean differences were specified, and even then there was little difference between empirical power of the two procedures. For nonstepwise mean differences, the Kruskal-Wallis test was found to be extremely sensitive. In fact, for small theoretical alphas, the power of the Kruskal-Wallis test diminished almost to zero for negative relationships of sample sizes and variances for both distributions. An interesting point concerning this lack of power for the Kruskal-Wallis test for nonstepwise mean differences was it's consistency. The empirical power values were found to be consistently low across distributions, across variance patterns, and across sample size patterns for N = 28. Due to the lack of power of the Kruskal-Wallis test in these cases, a recommendation for use at small alphas, with medium size total samples cannot be given on the basis of this research. Further, most researchers have no idea about the type of mean differences present in their data, and to expect them to be stepwise in nature is unreasonable. Therefore, a researcher utilizing small unequal sample sizes suspecting inequality of variances could be almost totally without power if testing hypotheses at α=.01 or smaller using the Kruskal-Wallis test. The F-test was found to generally have less power with nonstepwise mean differences than for stepwise mean differences in the negatively related cases but power for the F-test was much higher than that of the Kruskal-Wallis test and showed no indication of diminishing to zero. It should be noted, however, that for the larger unequal sample sizes investigated (N≐68), the Kruskal-Wallis test was competitive in terms of power for nonstepwise mean differences if the distribution was exponential. But the lack of power was still evident for the Kruskal-Wallis test even for larger sample sizes for normal distributions.

In conclusion, when normality and/or homogeneity of variance is doubtful, the ANOVA F-test was found to be the recommended procedure for testing hypotheses, especially at significance levels of .01 or less with a medium size sample. This recommendation is based primarily on the instability of power for the Kruskal-Wallis test. If a researcher is unconcerned with power or has a large total sample size, the Kruskal-Wallis test could be recommended as being competitive to the F-test when assumptions of normality and homogeneity of variance are violated.

REFERENCES

BOX, G. E. P. & ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departure from assumptions. J. R. Statis. Soc. B. 17, 1-34.

BOX, G. E. P. & MULLER, M. E. 1958). A note on the generation of random normal deviates. Ann. math. Statis. 29, 610-611.

BRADLEY, J. V. (1966). Studies in Research Methodology IV: A sampling Study of the Central Limit Theorem and the Robustness of One-Sample Parametric Tests. AMRL Technical Document Report, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

BRADLEY, J.V. (1966). Studies in Research Methodology VII: The Central Limit Effect for Two Dozen Populations and its Correlation With Population Moments. Behavioral Science Laboratoty, Wright-Patterson Air Force Base, Ohio.

BRADLEY, J. V. (1968). Distribution-Free Statistical Tests. London: Prentice-Hall.

FEIR, B & TOOTHAKER, L. E. (1974). An empirical comparison of the ANOVA F-test, the Kruskal-Wallis test, and the normal scores test under violation of assumptions. In Press. Educ. Psy. Meas.

HODGES, J. L. & LEHMAN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. Proc. Fourth Berkely Symp. Math. Statist. & Prob. Berkely and Los Angeles: University of California Press.

INTERNATIONAL BUSINESS MACHINES CORPORATION (1968). System/360, Scientific Subroutine Package. Programmer's Manual H20-0205-3, 77.

KLOTZ, J. (1963). Small sample power and efficiency for the one sample Wilcoxon and normal scores test. Ann. math. Statis. 34, 624-632.

KRUSKAL, W. H. (1952). A nonparametric test for the several sample problem. Ann. math. Statist. 23, 525-540.

LEHMAN, R. S. & BAILEY, D. E. (1968). Digital Computing. New York: John Wiley.

McSWEENEY, M. (1967). An empirical study of two proposed nonparametric tests for main effects and interaction. (Doctor's thesis, University of California, Berkeley.)

PEARSON, E. S. & HARTLEY, H. O. (1958). Chart of the power function for analysis of variance tests, derived from the non-central F-distribution. Biometrika 38. 112-130.

PENFIELD, D. & McSWEENY, M. E. (1968). The normal scores test for the two-sample problem. Psy. Bull. 69, 183-191.

PURI, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. Ann. math. Statist. 35, 102-121.

TERRY, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. Ann. math. Statis. 23, 346-366.

VAN DER LAAN, P. (1964). Exact power of some rank tests. Publ. Inst. Statist. Univ. Paris 13, 211-234.

TABLE 1

Values of n_j and μ_j

j	n_j	$\mu_{j_{SL}}$	$\mu_{j_{NSL}}$	$\mu_{j_{SS}}$	$\mu_{j_{NSS}}$
1	7	0	.577	0	.419
2	7	.67	.577	.487	.419
3	7	1.34	.577	.974	.419
4	7	2.01	2.308	1.461	1.676
1	4	0	.521	0	.379
2	6	.707	.521	.513	.379
3	8	1.414	.521	1.026	.379
4	10	2.121	2.084	1.539	1.516
1	4	0	.553	0	.402
2	8	.727	.553	.529	.402
3	8	1.454	.553	1.058	.402
4	8	2.181	2.212	1.587	1.608
1	6	0	.521	0	.379
2	6	.646	.521	.469	.379
3	6	1.292	.521	.938	.379
4	10	1.938	2.084	1.407	1.516
1	11	0	.315	0	.236
2	15	.415	.315	.310	.236
3	19	.830	.315	.620	.236
4	23	1.245	1.260	.930	.944
1	11	0	.332	0	.249
2	19	.426	.332	.319	.249
3	19	.852	.332	.638	.249
4	19	1.278	1.328	.957	.996

n_j = number of obwervations per treatment level

$\mu_{j_{SL}}$ = a stepwise difference such that power was approximately .86 for α=.05 for the normal distribution

$\mu_{j_{NSL}}$ = a non-stepwise difference (1:1:1:4) such that power was approximately .86 for α=.05 for the normal distribution

$\mu_{j_{SS}}$ = a stepwise difference such that power was approximately .60 for α=.05 for the normal distribution

$\mu_{j_{NSS}}$ = a non-stepwise difference (1:1:1:4) such that power was approximately .60 for α=.05 for the normal distribution

TABLE 2

Values of Variances

j	$=\sigma^2$	Stepwise	SSSL*	SLLL**
1	1.0	.4	.5714	.30769
2	1.0	.8	.5714	1.23076
3	1.0	1.2	.5714	1.23076
4	1.0	1.6	2.2857	1.23076

* SSSL - These variances are approximately in the ratio 1:1:1:4

** SLLL - These variances are approximately in the ratio 1:4:4:4

TABLE 3

Normal Population-Equal Sample Sizes (7,7,7,7)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$= \sigma_e^2$	α	.096	.043	.019	.007	.002	.090	.035	.014	.004	.001
	$1-\beta_{SL}$	.940	.872	.765	.640	.526	.921	.817	.702	.505	.361
	$1-\beta_{NSL}$	.936	.875	.803	.662	.569	.905	.832	.703	.480	.313
	$1-\beta_{SS}$	.719	.601	.485	.325	.234	.676	.538	.404	.226	.121
	$1-\beta_{NSS}$	.749	.613	.477	.331	.233	.697	.546	.376	.196	.104
$\neq \sigma_e^2$ (stepwise)	α	.104	.063	.042	.018	.009	.111	.054	.024	.005	.004
	$1-\beta_{SL}$	.957	.900	.814	.675	.579	.951	.880	.754	.582	.421
	$1-\beta_{NSL}$	.899	.829	.756	.645	.552	.831	.713	.594	.425	.275
	$1-\beta_{SS}$	.719	.590	.459	.322	.241	.704	.547	.401	.240	.150
	$1-\beta_{NSS}$	.678	.551	.429	.301	.231	.604	.430	.302	.172	.102
$\neq \sigma_e^2$(SSSL)	α	.103	.056	.028	.014	.009	.091	.035	.014	.007	.003
	$1-\beta_{SL}$	.921	.855	.774	.668	.557	.923	.836	.726	.536	.419
	$1-\beta_{NSL}$	.856	.798	.719	.601	.526	.753	.642	.515	.339	.236
	$1-\beta_{SS}$	.693	.576	.449	.318	.258	.692	.540	.384	.232	.145
	$1-\beta_{NSS}$	.645	.555	.459	.350	.276	.536	.397	.260	.147	.084
$\neq \sigma_e^2$(SLLL)	α	.098	.060	.025	.007	.004	.101	.051	.017	.007	.003
	$1-\beta_{SL}$	.962	.910	.823	.684	.578	.958	.907	.794	.593	.437
	$1-\beta_{NSL}$	.931	.860	.781	.678	.570	.898	.797	.682	.496	.334
	$1-\beta_{SS}$	.735	.588	.446	.292	.206	.758	.586	.406	.222	.122
	$1-\beta_{NSS}$	.709	.581	.455	.317	.247	.666	.504	.361	.191	.104

TABLE 4

Exponential Population-Equal Sample Sizes (7,7,7,7)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-Test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$= \sigma_e^2$	α	.094	.045	.017	.005	.000	.099	.052	.014	.002	.001
	$1-\beta$ SL	.919	.867	.793	.685	.616	.959	.910	.860	.715	.599
	$1-\beta$ NSL	.936	.895	.817	.722	.630	.980	.938	.852	.638	.427
	$1-\beta$ SS	.768	.643	.532	.397	.329	.855	.749	.615	.445	.315
	$1-\beta$ NSS	.744	.640	.531	.388	.290	.874	.739	.573	.341	.180
$\neq \sigma_e^2$ (stepwise)	α	.094	.054	.025	.009	.006	.109	.052	.024	.008	.004
	$1-\beta$ SL	.988	.956	.902	.821	.742	.995	.984	.963	.898	.826
	$1-\beta$ NSL	.961	.931	.878	.794	.712	.991	.967	.886	.696	.501
	$1-\beta$ SS	.878	.768	.638	.453	.350	.950	.879	.790	.607	.474
	$1-\beta$ NSS	.804	.698	.583	.446	.355	.890	.777	.615	.370	.200
$\neq \sigma_e^2$ SSSL	α	.116	.059	.026	.010	.009	.104	.054	.024	.006	.004
	$1-\beta$ SL	.974	.951	.906	.838	.761	.988	.974	.946	.872	.800
	$1-\beta$ NSL	.986	.965	.924	.840	.764	.997	.977	.915	.763	.544
	$1-\beta$ SS	.849	.748	.637	.488	.392	.915	.834	.740	.588	.453
	$1-\beta$ NSS	.860	.763	.658	.482	.375	.899	.797	.636	.359	.215
$\neq \sigma_e^2$ SLLL	α	.092	.047	.023	.010	.008	.103	.052	.022	.009	.004
	$1-\beta$ SL	.976	.959	.916	.832	.752	.997	.988	.966	.919	.846
	$1-\beta$ NSL	.946	.894	.822	.725	.640	.985	.951	.850	.645	.439
	$1-\beta$ SS	.884	.794	.671	.517	.389	.967	.918	.834	.656	.512
	$1-\beta$ NSS	.770	.671	.558	.429	.338	.903	.781	.612	.372	.203

TABLE 5

Normal Population-Unequal Sample Sizes, (4, 6, 8, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.127	.052	.030	.012	.007	.126	.050	.022	.005	.001
$= \sigma_e^2$	$1-\beta_{SL}$	.927	.860	.791	.652	.545	.906	.821	.685	.488	.341
	$1-\beta_{NSL}$	.922	.859	.781	.651	.559	.909	.833	.718	.545	.399
	$1-\beta_{SS}$	.724	.580	.464	.318	.239	.680	.523	.382	.199	.100
	$1-\beta_{NSS}$	.732	.600	.467	.314	.231	.699	.538	.384	.220	.118
$\neq$ n's	α	.078	.040	.017	.005	.003	.081	.032	.011	.003	.002
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.931	.865	.780	.644	.512	.933	.840	.706	.490	.314
	$1-\beta_{NSL}$	.901	.828	.738	.597	.501	.876	.786	.664	.477	.339
	$1-\beta_{SS}$	.691	.529	.401	.260	.183	.670	.490	.338	.179	.089
positively related	$1-\beta_{NSS}$	.655	.530	.404	.276	.207	.622	.468	.329	.199	.118
$\neq$ n's	α	.163	.082	.048	.023	.010	.112	.057	.028	.004	.001
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.944	.906	.849	.752	.672	.920	.860	.759	.597	.443
	$1-\beta_{NSL}$	.745	.636	.549	.441	.363	.595	.460	.308	.112	.047
	$1-\beta_{SS}$	.802	.675	.575	.429	.337	.747	.593	.450	.262	.154
negatively related	$1-\beta_{NSS}$	.541	.430	.335	.228	.168	.412	.277	.156	.050	.023
$\neq$ n's	α	.077	.041	.017	.008	.005	.078	.036	.007	.002	.000
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.906	.841	.759	.618	.517	.897	.807	.688	.497	.353
	$1-\beta_{NSL}$	.906	.829	.740	.613	.527	.872	.778	.653	.476	.337
	$1-\beta_{SS}$	.699	.552	.426	.284	.208	.665	.510	.348	.185	.105
positively related	$1-\beta_{NSS}$	.676	.556	.440	.291	.215	.632	.492	.343	.187	.113
$\neq$ n's	α	.188	.120	.085	.050	.038	.126	.076	.034	.007	.002
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.967	.928	.874	.782	.706	.953	.884	.806	.662	.512
	$1-\beta_{NSL}$	.702	.616	.527	.455	.379	.539	.403	.271	.085	.026
	$1-\beta_{SS}$	.798	.681	.571	.469	.377	.752	.607	.459	.291	.193
negatively related	$1-\beta_{NSS}$	.509	.421	.338	.264	.201	.356	.261	.162	.055	.016
$\neq$ n's	α	.099	.045	.025	.008	.007	.088	.038	.015	.005	.003
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.937	.868	.781	.639	.530	.919	.846	.700	.498	.325
	$1-\beta_{NSL}$	.920	.863	.783	.658	.550	.904	.820	.716	.532	.397
	$1-\beta_{SS}$	.679	.531	.412	.255	.166	.668	.503	.342	.173	.101
positively related	$1-\beta_{NSS}$	.680	.550	.421	.280	.189	.660	.497	.350	.180	.100
$\neq$ n's	α	.148	.079	.058	.024	.015	.122	.059	.026	.010	.002
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.966	.924	.875	.772	.698	.953	.905	.819	.625	.481
	$1-\beta_{NSL}$	.756	.638	.539	.413	.318	.677	.502	.315	.109	.042
	$1-\beta_{SS}$	.767	.652	.547	.395	.304	.752	.601	.443	.255	.158
negatively related	$1-\beta_{NSS}$	.507	.374	.272	.167	.122	.423	.270	.142	.053	.024

TABLE 6

Exponential Population-Unequal Sample Sizes, (4, 6, 8, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.094	.048	.029	.017	.010	.101	.043	.016	.010	.004
$\neq$ n's	$1-\beta$ SL	.941	.876	.811	.708	.627	.957	.910	.840	.732	.617
	$1-\beta$ NSL	.941	.899	.824	.701	.605	.974	.949	.888	.737	.581
$= \sigma_e^2$	$1-\beta$ SS	.745	.626	.538	.384	.295	.846	.738	.634	.458	.332
	$1-\beta$ NSS	.768	.640	.544	.394	.309	.886	.796	.646	.465	.302
$\neq$ n's	α	.075	.029	.020	.010	.006	.081	.034	.015	.006	.002
$\neq \sigma^2$ (stepwise)	$1-\beta$ SL	.946	.902	.847	.739	.644	.981	.951	.912	.808	.675
	$1-\beta$ NSL	.921	.871	.798	.682	.609	.977	.934	.8[illegible]7	.739	.579
	$1-\beta$ SS	.799	.662	.547	.388	.294	.886	.800	.694	.537	.370
positively related	$1-\beta$ NSS	.780	.663	.541	.394	.300	.887	.798	.665	.463	.299
= n's	α	.165	.105	.054	.032	.019	.156	.084	.043	.012	.004
$\neq \sigma^2$ (stepwise)	$1-\beta$ SL	.991	.976	.949	.902	.847	.997	.992	.975	.925	.858
	$1-\beta$ NSL	.854	.763	.675	.562	.464	.866	.650	.382	.131	.055
	$1-\beta$ SS	.911	.832	.739	.611	.520	.959	.894	.828	.639	.480
negatively related	$1-\beta$ NSS	.631	.517	.418	.295	.222	.619	.407	.229	.087	.029
= n's	α	.064	.036	.019	.010	.003	.075	.035	.016	.001	.000
	$1-\beta$ SL	.964	.921	.867	.777	.688	.979	.959	.917	.831	.722
$\neq \sigma_e^2$ (SSSL)	$1-\beta$ NSL	.968	.931	.872	.775	.687	.990	.971	.917	.816	.680
	$1-\beta$ SS	.796	.670	.561	.404	.306	.878	.788	.675	.505	.385
positively related	$1-\beta$ NSS	.808	.714	.574	.402	.305	.914	.823	.707	.494	.344
$\neq$ n's	α	.179	.103	.059	.031	.020	.130	.070	.043	.012	.006
$\neq \sigma^2$ (SSSL)	$1-\beta$ SL	.976	.956	.924	.872	.809	.991	.977	.954	.886	.788
	$1-\beta$ NSL	.871	.803	.713	.583	.507	.873	.660	.393	.131	.045
	$1-\beta$ SS	.876	.814	.736	.607	.506	.921	.857	.761	.588	.446
negatively related	$1-\beta$ NSS	.670	.549	.439	.318	.235	.583	.341	.159	.054	.018
$\neq$ n's	α	.071	.030	.016	.010	.005	.083	.033	.017	.009	.004
$\neq \sigma^2$ (SLLL)	$1-\beta$ SL	.960	.916	.848	.729	.638	.976	.954	.909	.801	.678
	$1-\beta$ NSL	.944	.889	.818	.716	.600	.983	.945	.886	.732	.586
	$1-\beta$ SS	.798	.682	.548	.409	.304	.902	.809	.712	.526	.400
positively related	$1-\beta$ NSS	.745	.628	.487	.350	.282	.885	.774	.626	.408	.264
$\neq$ n's	α	.140	.083	.046	.021	.015	.142	.082	.034	.011	.002
$\neq \sigma^2$ (SLLL)	$1-\beta$ SL	.987	.965	.929	.869	.799	1.000	.995	.979	.913	.843
	$1-\beta$ NSL	.825	.731	.630	.510	.408	.875	.663	.399	.157	.077
	$1-\beta$ SS	.913	.843	.727	.595	.466	.970	.919	.841	.645	.463
negatively related	$1-\beta$ NSS	.622	.502	.388	.273	.210	.659	.412	.232	.086	.038

TABLE 7

Normal Population-Unequal Sample Sizes, (4, 8, 8, 8)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large;NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.092	.042	.021	.008	.005	.090	.037	.016	.006	.002
$= \sigma_e^2$	$1-\beta_{SL}$	.935	.865	.788	.664	.566	.907	.836	.702	.518	.353
	$1-\beta_{NSL}$	.937	.878	.798	.675	.562	.921	.838	.705	.516	.359
	$1-\beta_{SS}$	.745	.620	.484	.342	.244	.703	.557	.398	.209	.117
	$1-\beta_{NSS}$	.706	.574	.465	.312	.228	.677	.530	.364	.198	.111
$\neq$ n's	α	.086	.042	.025	.006	.004	.082	.034	.014	.007	.003
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.945	.878	.798	.658	.548	.919	.842	.725	.511	.351
	$1-\beta_{NSL}$	.946	.874	.794	.669	.553	.915	.820	.687	.503	.346
	$1-\beta_{SS}$	.709	.573	.456	.300	.220	.681	.516	.355	.193	.109
positively related	$1-\beta_{NSS}$	.708	.575	.465	.327	.234	.654	.519	.365	.207	.125
$\neq$ n's	α	.124	.061	.031	.013	.011	.105	.047	.018	.007	.002
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.956	.903	.835	.708	.616	.945	.855	.760	.567	.409
	$1-\beta_{NSL}$	.769	.670	.553	.418	.303	.655	.489	.313	.090	.028
	$1-\beta_{SS}$	.781	.667	.542	.384	.298	.747	.594	.437	.270	.167
negatively related	$1-\beta_{NSS}$	.554	.420	.319	.208	.142	.461	.295	.157	.045	.014
$\neq$ n's	α	.106	.054	.027	.011	.004	.111	.046	.020	.003	.001
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.949	.889	.817	.706	.609	.932	.848	.747	.545	.401
	$1-\beta_{NSL}$	.916	.853	.778	.666	.568	.881	.789	.662	.483	.328
	$1-\beta_{SS}$	.713	.577	.463	.329	.234	.671	.514	.382	.210	.129
positively related	$1-\beta_{NSS}$	.709	.576	.455	.309	.230	.640	.495	.345	.189	.106
$\neq$ n's	α	.126	.070	.043	.024	.013	.094	.050	.021	.003	.000
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.951	.911	.852	.742	.659	.935	.859	.764	.592	.440
	$1-\beta_{NSL}$	.784	.699	.592	.471	.388	.636	.501	.321	.097	.030
	$1-\beta_{SS}$	.782	.677	.553	.414	.314	.732	.589	.444	.257	.151
negatively related	$1-\beta_{NSS}$	.563	.448	.351	.251	.179	.455	.290	.163	.050	.018
$\neq$ n's	α	.080	.042	.021	.010	.004	.080	.033	.011	.005	.000
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.938	.891	.806	.685	.571	.933	.849	.722	.541	.371
	$1-\beta_{NSL}$	.935	.870	.783	.663	.573	.917	.809	.699	.524	.358
	$1-\beta_{SS}$	.711	.563	.427	.261	.196	.684	.526	.346	.178	.092
positively related	$1-\beta_{NSS}$	.691	.555	.425	.295	.218	.664	.497	.341	.184	.098
$\neq$ n's	α	.092	.049	.029	.007	.004	.092	.042	.018	.006	.002
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.966	.916	.837	.711	.600	.960	.897	.768	.568	.414
	$1-\beta_{NSL}$	.788	.656	.543	.408	.315	.702	.512	.311	.101	.039
	$1-\beta_{SS}$	.784	.671	.560	.383	.260	.766	.629	.469	.236	.125
negatively related	$1-\beta_{NSS}$	.556	.408	.289	.172	.120	.474	.291	.158	.044	.015

TABLE 8

Exponential Population-Unequal Sample Sizes, (4, 8, 8, 8)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL - nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.01	.05	.025	.01	.005	.01	.05	.025	.01	.005
≠ n's	α	.095	.051	.026	.012	.008	.095	.042	.020	.004	.000
$= \sigma_e^2$	$1-\beta_{SL}$	.922	.868	.809	.712	.607	.961	.922	.854	.729	.584
	$1-\beta_{NSL}$	.950	.897	.842	.722	.628	.985	.948	.865	.684	.500
	$1-\beta_{SS}$	.776	.655	.545	.415	.323	.874	.774	.664	.478	.340
	$1-\beta_{NSS}$	.766	.650	.533	.400	.309	.887	.762	.626	.378	.215
≠ n's	α	.067	.037	.020	.008	.006	.076	.033	.015	.005	.002
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.938	.887	.817	.709	.625	.970	.932	.878	.768	.649
	$1-\beta_{NSL}$	.938	.880	.805	.714	.640	.981	.941	.856	.670	.497
positively related	$1-\beta_{SS}$	.775	.653	.546	.386	.286	.883	.794	.683	.478	.332
	$1-\beta_{NSS}$	.766	.664	.561	.424	.305	.886	.764	.620	.418	.250
≠ n's	α	.099	.055	.035	.017	.011	.122	.048	.017	.004	.003
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{SL}$	.968	.935	.899	.806	.728	.992	.973	.931	.833	.732
	$1-\beta_{NSL}$	.830	.741	.633	.494	.418	.866	.632	.329	.096	.041
negatively related	$1-\beta_{SS}$	.816	.728	.619	.488	.384	.907	.809	.704	.534	.386
	$1-\beta_{NSS}$	.590	.449	.339	.230	.167	.591	.345	.178	.051	.017
≠ n's	α	.078	.034	.018	.007	.005	.090	.036	.013	.005	.001
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.955	.914	.850	.747	.663	.975	.940	.886	.774	.640
	$1-\beta_{NSL}$	.958	.911	.848	.742	.671	.981	.954	.883	.722	.558
positively related	$1-\beta_{SS}$	.796	.675	.561	.438	.333	.881	.795	.675	.496	.349
	$1-\beta_{NSS}$	.787	.662	.543	.408	.310	.890	.796	.646	.394	.234
≠ n's	α	.111	.056	.025	.007	.005	.101	.051	.021	.004	.002
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{SL}$	.968	.930	.878	.798	.730	.988	.967	.926	.820	.703
	$1-\beta_{NSL}$	.875	.781	.701	.571	.464	.890	.671	.394	.118	.036
negatively related	$1-\beta_{SS}$	.834	.764	.681	.538	.454	.907	.829	.713	.541	.413
	$1-\beta_{NSS}$	.639	.523	.420	.301	.234	.641	.402	.205	.058	.020
≠ n's	α	.074	.042	.020	.010	.006	.083	.033	.015	.007	.002
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.941	.900	.827	.732	.655	.978	.949	.909	.790	.652
	$1-\beta_{NSL}$	.942	.889	.818	.706	.622	.983	.940	.851	.677	.513
positively related	$1-\beta_{SS}$	.770	.661	.549	.408	.298	.887	.783	.673	.479	.357
	$1-\beta_{NSS}$	.757	.658	.527	.386	.284	.904	.775	.616	.362	.214
≠ n's	α	.088	.047	.021	.015	.010	.095	.051	.024	.007	.003
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{SL}$	.965	.932	.878	.777	.702	.992	.971	.935	.831	.726
	$1-\beta_{NSL}$	.812	.708	.608	.493	.387	.836	.624	.316	.100	.041
negatively related	$1-\beta_{SS}$	.845	.740	.620	.479	.375	.928	.837	.732	.552	.401
	$1-\beta_{NSS}$	.587	.471	.366	.254	.192	.633	.386	.211	.068	.021

TABLE 9

Normal Population-Unequal Sample Sizes, (6, 6, 6, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and The Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($= \sigma_e^2$; $\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.113	.061	.032	.010	.009	.095	.053	.020	.006	.001
$= \sigma^2$	$1-\beta_{SL}$	.948	.900	.824	.679	.575	.936	.853	.743	.540	.399
	$1-\beta_{NSL}$	.940	.873	.790	.680	.564	.921	.835	.736	.546	.388
	$1-\beta_{SS}$	.724	.603	.486	.327	.246	.696	.534	.387	.231	.143
	$1-\beta_{NSS}$	.739	.617	.479	.327	.240	.720	.552	.402	.232	.140
$\neq$ n's	α	.075	.035	.020	.012	.005	.071	.036	.015	.003	.001
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.936	.857	.745	.596	.486	.928	.842	.722	.524	.374
positively related	$1-\beta_{NSL}$	.895	.800	.701	.566	.482	.840	.737	.603	.448	.320
	$1-\beta_{SS}$	.696	.565	.417	.287	.196	.706	.541	.379	.225	.129
	$1-\beta_{NSS}$	.665	.521	.395	.258	.190	.607	.455	.308	.168	.107
$\neq$ n's	α	.109	.052	.031	.010	.003	.102	.045	.018	.003	.000
$\neq \sigma^2$ (stepwise)	$1-\beta_{SL}$	.960	.904	.812	.683	.588	.939	.847	.727	.550	.417
negatively related	$1-\beta_{NSL}$	.853	.769	.665	.525	.414	.797	.668	.518	.303	.163
	$1-\beta_{SS}$	.762	.646	.518	.366	.268	.739	.583	.443	.240	.140
	$1-\beta_{NSS}$	.606	.483	.361	.237	.175	.534	.389	.248	.116	.062
$\neq$ n's	α	.057	.033	.021	.009	.006	.069	.032	.011	.005	.002
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.852	.773	.632	.482	.396	.875	.771	.623	.442	.315
positively related	$1-\beta_{NSL}$	.831	.728	.636	.512	.427	.753	.631	.508	.341	.226
	$1-\beta_{SS}$	.559	.444	.327	.215	.156	.607	.430	.304	.167	.106
	$1-\beta_{NSS}$	.569	.445	.337	.235	.175	.506	.364	.257	.153	.099
$\neq$ n's	α	.103	.064	.030	.012	.007	.105	.058	.022	.005	.003
$\neq \sigma^2$ (SSSL)	$1-\beta_{SL}$	.940	.873	.812	.693	.589	.916	.836	.732	.556	.421
negatively related	$1-\beta_{NSL}$	.832	.758	.652	.500	.424	.771	.642	.495	.300	.156
	$1-\beta_{SS}$	.753	.626	.501	.367	.271	.707	.547	.402	.249	.147
	$1-\beta_{NSS}$	.638	.496	.382	.253	.173	.552	.394	.251	.121	.066
$\neq$ n's	α	.102	.047	.021	.009	.006	.087	.032	.019	.005	.001
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.942	.894	.803	.641	.527	.961	.900	.787	.568	.378
positively related	$1-\beta_{NSL}$	.935	.872	.795	.644	.532	.917	.843	.713	.507	.351
	$1-\beta_{SS}$	.709	.576	.432	.293	.206	.744	.596	.427	.238	.135
	$1-\beta_{NSS}$	.693	.557	.432	.301	.219	.669	.517	.352	.201	.125
$\neq$ n's	α	.101	.048	.028	.011	.007	.089	.036	.012	.007	.003
$\neq \sigma^2$ (SLLL)	$1-\beta_{SL}$	.945	.882	.831	.705	.591	.923	.854	.748	.557	.410
negatively related	$1-\beta_{NSL}$	.875	.795	.693	.540	.439	.844	.698	.540	.335	.181
	$1-\beta_{SS}$	.746	.622	.498	.363	.256	.729	.580	.420	.233	.135
	$1-\beta_{NSS}$	.641	.490	.363	.237	.172	.577	.400	.251	.137	.065

TABLE 10

Exponential Population-Unequal Sample Sizes, (6, 6, 6, 10)

[Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under four variance cases ($=\sigma_e^2$; $\neq\sigma_e^2$, stepwise; $\neq\sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq\sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4)]

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
$\neq$ n's	α	.088	.035	.019	.006	.005	.100	.042	.021	.006	.002
$=\sigma^2$	$1-\beta_{SL}$	.923	.875	.813	.710	.635	.961	.929	.878	.751	.642
	$1-\beta_{NSL}$	.942	.893	.820	.701	.603	.980	.943	.871	.705	.563
	$1-\beta_{SS}$	.757	.646	.557	.393	.307	.856	.771	.655	.481	.350
	$1-\beta_{NSS}$	.763	.651	.551	.392	.316	.908	.806	.648	.453	.319
$\neq$ n's	α	.071	.036	.017	.006	.002	.073	.031	.013	.005	.004
	$1-\beta_{SL}$	.956	.896	.836	.747	.668	.984	.951	.887	.784	.672
$\neq\sigma^2$ (stepwise)	$1-\beta_{NSL}$	.936	.883	.805	.694	.600	.979	.947	.888	.727	.584
	$1-\beta_{SS}$	.756	.648	.537	.416	.314	.861	.765	.647	.491	.373
positively related	$1-\beta_{NSS}$	.775	.643	.539	.384	.297	.892	.785	.650	.445	.294
$\neq$ n's	α	.096	.054	.031	.015	.005	.098	.039	.017	.006	.002
$\neq\sigma^2$ (stepwise)	$1-\beta_{SL}$	.955	.915	.864	.770	.672	.986	.963	.903	.802	.679
	$1-\beta_{NSL}$	.892	.813	.717	.594	.503	.943	.855	.675	.410	.217
negatively related	$1-\beta_{SS}$	.806	.714	.597	.458	.375	.896	.819	.701	.509	.359
	$1-\beta_{NSS}$	.678	.546	.433	.289	.210	.777	.608	.398	.187	.067
$\neq$ n's	α	.081	.040	.020	.012	.006	.085	.036	.012	.005	.000
$\neq\sigma^2$ (SSSL)	$1-\beta_{SL}$	.950	.898	.834	.740	.650	.978	.945	.879	.770	.657
	$1-\beta_{NSL}$	.957	.905	.833	.724	.624	.990	.950	.890	.748	.596
positively related	$1-\beta_{SS}$	.778	.677	.572	.424	.316	.873	.786	.658	.504	.386
	$1-\beta_{NSS}$	.793	.661	.546	.398	.316	.886	.796	.677	.466	.310
$\neq$ n's	α	.097	.052	.024	.011	.005	.104	.050	.025	.011	.003
$\neq\sigma^2$ (SSSL)	$1-\beta_{SL}$	.946	.907	.844	.750	.683	.982	.949	.892	.781	.671
	$1-\beta_{NSL}$	.875	.800	.707	.577	.497	.936	.841	.673	.387	.182
negatively related	$1-\beta_{SS}$	.800	.717	.598	.433	.337	.897	.805	.669	.487	.334
	$1-\beta_{NSS}$	.671	.537	.424	.309	.232	.775	.599	.394	.168	.070
$\neq$ n's	α	.077	.048	.024	.009	.005	.089	.037	.012	.003	.000
$\neq\sigma^2$ (SLLL)	$1-\beta_{SL}$	.933	.890	.820	.735	.659	.975	.930	.872	.771	.673
	$1-\beta_{NSL}$	.928	.862	.792	.686	.601	.972	.931	.870	.719	.571
positively related	$1-\beta_{SS}$	.753	.641	.521	.387	.292	.862	.757	.650	.482	.352
	$1-\beta_{NSS}$	.782	.657	.547	.421	.315	.897	.808	.671	.461	.320
$\neq$ n's	α	.102	.045	.028	.016	.005	.106	.048	.023	.007	.002
$\neq\sigma^2$ (SLLL)	$1-\beta_{SL}$	.950	.907	.848	.755	.678	.985	.957	.908	.782	.654
	$1-\beta_{NSL}$	.884	.797	.714	.589	.503	.945	.859	.678	.408	.207
negatively related	$1-\beta_{SS}$	.831	.724	.599	.451	.348	.911	.803	.673	.501	.368
	$1-\beta_{NSS}$	.687	.578	.479	.348	.241	.790	.634	.449	.190	.090

TABLE 11

Normal Population-Unequal Sample Sizes

(Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under three variance cases ($\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4).)

Sample Sizes = 11, 15, 19, 23

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.074	.039	.017	.005	.002	.079	.037	.018	.005	.002
$\neq$ n's	$1-\beta_{SL}$	.915	.819	.715	.586	.484	.917	.844	.755	.598	.487
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.877	.803	.706	.570	.487	.849	.743	.623	.487	.384
positively related	$1-\beta_{SS}$	.666	.535	.424	.296	.215	.692	.545	.438	.294	.206
	$1-\beta_{NSS}$	.652	.529	.409	.278	.211	.617	.481	.332	.233	.158
	α	.123	.064	.033	.014	.005	.107	.044	.027	.007	.003
$\neq$ n's	$1-\beta_{SL}$	.941	.895	.828	.730	.639	.925	.865	.788	.660	.542
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.774	.676	.568	.467	.387	.706	.579	.468	.321	.239
negatively related	$1-\beta_{SS}$	.762	.657	.545	.411	.310	.720	.604	.478	.341	.237
	$1-\beta_{NSS}$	.530	.410	.322	.214	.175	.454	.329	.223	.135	.093
	α	.069	.036	.016	.010	.003	.076	.037	.012	.007	.004
$\neq$ n's	$1-\beta_{SL}$	.930	.853	.761	.641	.543	.948	.894	.804	.681	.560
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.883	.804	.728	.612	.534	.880	.795	.690	.557	.455
positively related	$1-\beta_{SS}$	.703	.549	.430	.269	.182	.750	.611	.465	.291	.198
	$1-\beta_{NSS}$	.650	.529	.412	.301	.210	.629	.490	.382	.252	.174
	α	.157	.083	.047	.024	.014	.138	.065	.033	.015	.010
$\neq$ n's	$1-\beta_{SL}$	.964	.928	.870	.778	.685	.959	.927	.871	.747	.633
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.757	.662	.574	.453	.358	.710	.587	.473	.346	.248
negatively related	$1-\beta_{SS}$	.791	.666	.552	.419	.323	.783	.667	.539	.393	.298
	$1-\beta_{NSS}$	.546	.433	.336	.237	.174	.499	.360	.259	.160	.115
	α	.073	.041	.021	.011	.005	.090	.039	.023	.006	.002
$\neq$ n's	$1-\beta_{SL}$	.854	.776	.701	.575	.487	.885	.808	.699	.573	.476
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.792	.704	.624	.516	.442	.738	.617	.505	.401	.311
positively related	$1-\beta_{SS}$	.625	.498	.370	.268	.201	.661	.502	.401	.275	.207
	$1-\beta_{NSS}$	.553	.442	.353	.261	.202	.507	.386	.274	.189	.128
	α	.155	.087	.062	.032	.014	.111	.061	.029	.012	.005
$\neq$ n's	$1-\beta_{SL}$	.945	.899	.843	.747	.663	.943	.894	.811	.690	.597
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.709	.642	.574	.500	.429	.562	.450	.353	.234	.164
negativelyrelated	$1-\beta_{SS}$	.760	.652	.558	.436	.352	.727	.600	.495	.363	.261
	$1-\beta_{NSS}$	.529	.439	.373	.300	.247	.380	.274	.204	.130	.085

TABLE 12

Exponential Population-Unequal Sample Size

(Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise small). Alpha and power are observed under three variance cases ($\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4).)

Sample Sizes = 11, 15, 19, 23

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.079	.036	.012	.007	.005	.069	.033	.016	.010	.007
$\neq$ n's	$1-\beta_{SL}$	.983	.964	.913	.828	.731	.999	.993	.985	.964	.932
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.937	.889	.825	.732	.663	.994	.979	.957	.911	.863
positively related	$1-\beta_{SS}$	.865	.739	.619	.457	.355	.961	.914	.848	.748	.658
	$1-\beta_{NSS}$	.776	.667	.547	.406	.306	.928	.858	.750	.619	.504
	α	.122	.061	.033	.018	.010	.119	.061	.031	.013	.006
$\neq$ n's	$1-\beta_{SL}$	.982	.954	.918	.844	.781	.997	.992	.984	.963	.938
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.843	.745	.641	.499	.414	.970	.915	.823	.632	.491
negatively related	$1-\beta_{SS}$	.844	.762	.668	.531	.446	.965	.918	.840	.721	.620
	$1-\beta_{NSS}$	.611	.501	.402	.275	.224	.797	.650	.500	.302	.184
	α	.069	.033	.013	.003	.000	.092	.042	.021	.004	.003
$\neq$ n's	$1-\beta_{SL}$	.985	.956	.918	.822	.713	.998	.997	.996	.991	.970
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.937	.867	.797	.687	.597	.996	.990	.982	.943	.894
positively related	$1-\beta_{SS}$	.859	.738	.598	.434	.318	.971	.939	.896	.805	.713
	$1-\beta_{NSS}$	.777	.650	.527	.373	.287	.937	.891	.810	.690	.571
	α	.157	.086	.046	.024	.015	.140	.080	.041	.021	.007
$\neq$ n's	$1-\beta_{SL}$	.996	.990	.970	.932	.874	1.00	.999	.998	.989	.979
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.887	.785	.683	.571	.480	.984	.948	.879	.735	.565
negatively related	$1-\beta_{SS}$	.944	.891	.799	.653	.551	.990	.967	.937	.855	.776
	$1-\beta_{NSS}$	.674	.542	.423	.306	.243	.839	.710	.562	.387	.270
	α	.104	.053	.032	.016	.006	.116	.065	.042	.023	.011
$\neq$ n's	$1-\beta_{SL}$	.987	.962	.916	.825	.740	.995	.982	.964	.926	.881
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.969	.940	.894	.833	.753	.996	.975	.949	.896	.842
positively related	$1-\beta_{SS}$	.867	.768	.653	.507	.403	.919	.846	.770	.655	.566
	$1-\beta_{NSS}$	.868	.777	.661	.525	.419	.883	.793	.686	.539	.440
	α	.197	.123	.082	.048	.032	.144	.091	.053	.022	.010
$\neq$ n's	$1-\beta_{SL}$	.978	.963	.942	.909	.858	.996	.993	.983	.965	.947
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.910	.858	.787	.694	.613	.924	.861	.754	.606	.482
negatively related	$1-\beta_{SS}$	.877	.810	.739	.625	.528	.961	.909	.849	.738	.654
	$1-\beta_{NSS}$	.749	.660	.573	.466	.392	.706	.565	.432	.280	.183

TABLE 13

Normal Population-Unequal Sample Sizes

(Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise, small). Alpha and power are observed under three variance cases ($\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4).)

Sample Sizes = 11, 19, 19, 19

		F					KW				
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.084	.035	.021	.011	.007	.089	.041	.020	.007	.002
$\neq$ n's	$1-\beta_{SL}$	.908	.828	.748	.608	.503	.909	.832	.732	.586	.473
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.867	.781	.702	.583	.505	.806	.711	.600	.458	.369
positively related	$1-\beta_{SS}$	.709	.563	.435	.286	.213	.711	.561	.430	.279	.191
	$1-\beta_{NSS}$	.654	.519	.394	.297	.224	.588	.447	.324	.208	.153
	α	.154	.092	.050	.026	.017	.134	.068	.031	.016	.007
$\neq$ n's	$1-\beta_{SL}$	.942	.887	.829	.723	.640	.927	.870	.797	.673	.579
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.794	.692	.606	.495	.414	.696	.572	.442	.326	.231
negatively related	$1-\beta_{SS}$	.770	.649	.548	.408	.317	.748	.624	.498	.346	.240
	$1-\beta_{NSS}$	.581	.468	.378	.254	.195	.474	.356	.248	.143	.091
	α	.077	.040	.019	.006	.003	.082	.038	.016	.005	.003
$\neq$ n's	$1-\beta_{SL}$	.928	.856	.757	.601	.500	.943	.882	.793	.628	.517
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.877	.790	.719	.587	.508	.854	.767	.663	.538	.437
positively related	$1-\beta_{SS}$	.709	.580	.439	.283	.220	.753	.614	.472	.318	.226
	$1-\beta_{NSS}$	.667	.548	.441	.316	.241	.663	.529	.417	.274	.190
	α	.131	.070	.043	.025	.010	.127	.065	.036	.013	.007
$\neq$ n's	$1-\beta_{SL}$	.939	.879	.813	.695	.619	.947	.893	.822	.708	.605
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.775	.669	.565	.440	.357	.728	.603	.484	.339	.239
negatively related	$1-\beta_{SS}$	.748	.619	.502	.368	.262	.776	.646	.522	.356	.266
	$1-\beta_{NSS}$	.555	.448	.347	.235	.170	.526	.386	.264	.172	.101
	α	.099	.054	.026	.014	.011	.105	.053	.023	.012	.006
$\neq$ n's	$1-\beta_{SL}$	.872	.804	.708	.597	.513	.898	.797	.694	.571	.460
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.812	.756	.677	.573	.503	.733	.627	.507	.408	.324
positively related	$1-\beta_{SS}$	.676	.552	.441	.318	.224	.693	.553	.422	.281	.206
	$1-\beta_{NSS}$	.614	.501	.399	.301	.252	.510	.384	.282	.190	.141
	α	.207	.141	.093	.059	.037	.161	.080	.041	.015	.009
$\neq$ n's	$1-\beta_{SL}$	.937	.885	.832	.735	.662	.933	.890	.809	.706	.604
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.760	.689	.615	.522	.464	.607	.479	.381	.271	.202
negatively related	$1-\beta_{SS}$	.778	.672	.574	.453	.381	.758	.657	.538	.382	.291
	$1-\beta_{NSS}$	.582	.495	.409	.318	.261	.426	.315	.225	.136	.089

TABLE 14

Exponential Population-Unequal Sample Size

(Entries are the proportion of rejections in 1,000 experiments for the ANOVA F-test (F) and the Kruskal-Wallis Test (KW) in terms of probability of a Type I error (α) and power ($1-\beta$). Power is based on four cases of mean differences (SL-stepwise, large; NSL-nonstepwise, large; SS-stepwise, small; NSS-nonstepwise; small). Alpha and power are observed under three variance cases ($\neq \sigma_e^2$, stepwise; $\neq \sigma_e^2$, SSSL, approximately in the ratio of 1:1:1:4; and $\neq \sigma_e^2$, SLLL, approximately in the ratio of 1:4:4:4).)

Sample Sizes = 11, 19, 19, 19

				F					KW		
	α set	.10	.05	.025	.01	.005	.10	.05	.025	.01	.005
	α	.097	.049	.026	.010	.003	.094	.048	.029	.010	.003
$\neq$ n's	$1-\beta_{SL}$	.986	.963	.924	.845	.765	.999	.995	.991	.966	.947
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.949	.903	.836	.752	.671	.994	.984	.967	.917	.872
positively related	$1-\beta_{SS}$	.883	.773	.657	.521	.404	.959	.928	.877	.778	.679
	$1-\beta_{NSS}$	.792	.703	.589	.457	.376	.934	.866	.769	.631	.523
	α	.120	.064	.036	.018	.009	.127	.065	.037	.016	.007
$\neq$ n's	$1-\beta_{SL}$	.993	.979	.962	.908	.850	.999	.998	.992	.980	.956
$\neq \sigma_e^2$ (stepwise)	$1-\beta_{NSL}$	.882	.817	.720	.614	.527	.976	.931	.849	.705	.562
negatively related	$1-\beta_{SS}$	.922	.874	.794	.657	.561	.974	.951	.903	.822	.743
	$1-\beta_{NSS}$	.706	.598	.497	.368	.293	.822	.683	.534	.348	.230
	α	.082	.042	.020	.013	.007	.091	.042	.019	.011	.008
$\neq$ n's	$1-\beta_{SL}$	.973	.944	.893	.784	.697	1.00	.997	.994	.975	.954
$\neq \sigma_e^2$)SLLL)	$1-\beta_{NSL}$	.932	.868	.784	.675	.583	.994	.989	.974	.922	.862
positively related	$1-\beta_{SS}$	.851	.743	.617	.441	.343	.973	.941	.887	.799	.721
	$1-\beta_{NSS}$	.744	.638	.533	.396	.308	.945	.875	.795	.659	.561
	α	.123	.072	.044	.022	.013	.132	.072	.033	.013	.008
$\neq$ n's	$1-\beta_{SL}$	.994	.984	.957	.896	.826	1.00	.996	.993	.984	.972
$\neq \sigma_e^2$ (SLLL)	$1-\beta_{NSL}$	.883	.783	.686	.573	.470	.989	.959	.888	.726	.583
negatively related	$1-\beta_{SS}$	.927	.859	.740	.581	.485	.986	.966	.932	.842	.756
	$1-\beta_{NSS}$	.637	.523	.426	.314	.251	.836	.716	.579	.385	.272
	α	.120	.068	035	.016	.011	.116	.064	.034	.014	.006
$\neq$ n's	$1-\beta_{SL}$	.984	.961	.927	.855	.781	.995	.986	.973	.928	.890
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.983	.957	.922	.857	.810	.998	.988	.969	.934	.867
positively related	$1-\beta_{SS}$	.886	.810	.707	.573	.472	.933	.880	.802	.687	.602
	$1-\beta_{NSS}$	.877	.809	.732	.610	.521	.900	.817	.725	.583	.469
	α	.197	.124	.081	.044	.031	.150	.084	.040	.017	.009
$\neq$ n's	$1-\beta_{SL}$	.983	.975	.946	.894	.849	.996	.992	.980	.956	.929
$\neq \sigma_e^2$ (SSSL)	$1-\beta_{NSL}$	.938	.878	.837	.744	.672	.969	.919	.836	.677	.538
negatively related	$1-\beta_{SS}$	.895	.846	.775	.669	.572	.953	.919	.870	.775	.708
	$1-\beta_{NSS}$	.779	.681	.595	.498	.419	.743	.624	.498	.332	.221

TABLE 15

Schematic of Conditions Investigated ("X" indicates inclusion of the case. Each case was investigated for a normal population and an exponential population using both the ANOVA F-test and the Kruskal-Wallis Test).

		$\neq n_j$'s						
		(7,7,7,7)	(4,6,8,10)	(4,8,8,8)	(6,6,6,10)	(11,15,19,23)	(11,19,19,19)	
$= \sigma_e^2$	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X			
$\neq \sigma_e^2$ stepwise	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X	X	X	pos. related n_j's and σ_e^2's
			X	X	X	X	X	neg. related n_j's and σ_e^2's
$\neq \sigma_e^2$ SLLL	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X	X	X	pos. related n_j's and σ_e^2's
			X	X	X	X	X	neg. related n_j's and σ_e^2's
$\neq \sigma_e^2$ SSSL	α $1-\beta_{SL}$ $1-\beta_{NSL}$ $1-\beta_{SS}$ $1-\beta_{NSS}$	X	X	X	X	X	X	pos. related n_j's and σ_e^2's
			X	X	X	X	X	neg. related n_j's and σ_e^2's

n_j - sample size per treatment level

SL - a stepwise difference such that power was approximately .86 for α=.05 for the normal distribution

NSL - a nonstepwise difference (1:1:1:4) such that power was approximately .86 for α=.05 for the normal distribution

SS - a stepwise difference such that power was approximately .60 for α=.05 for the normal distribution

NSS - a nonstepwise difference (1:1:1:4) such that power was approximately .60 for α=.05 for the normal distribution

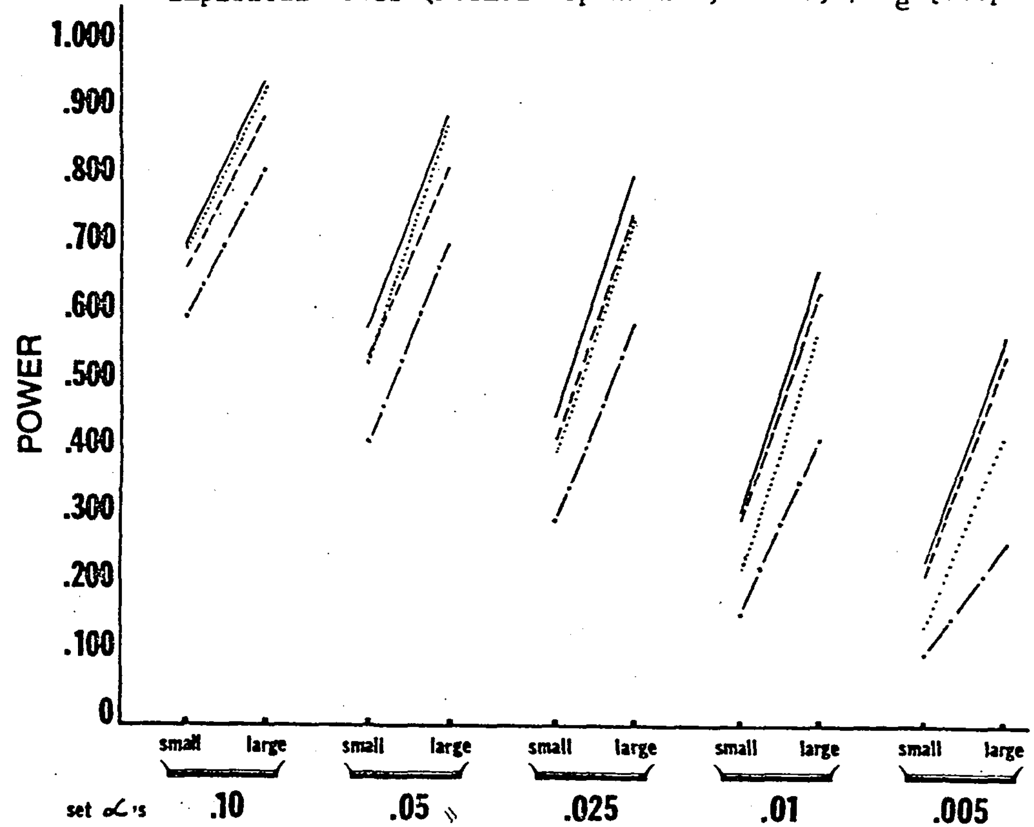
Stepwise - variances in the ratio of 1:2:3:4

SLLL - variances in the ratio of 1:4:4:4

SSSL - variances in the ratio of 1:1:1:4

FIGURE 1

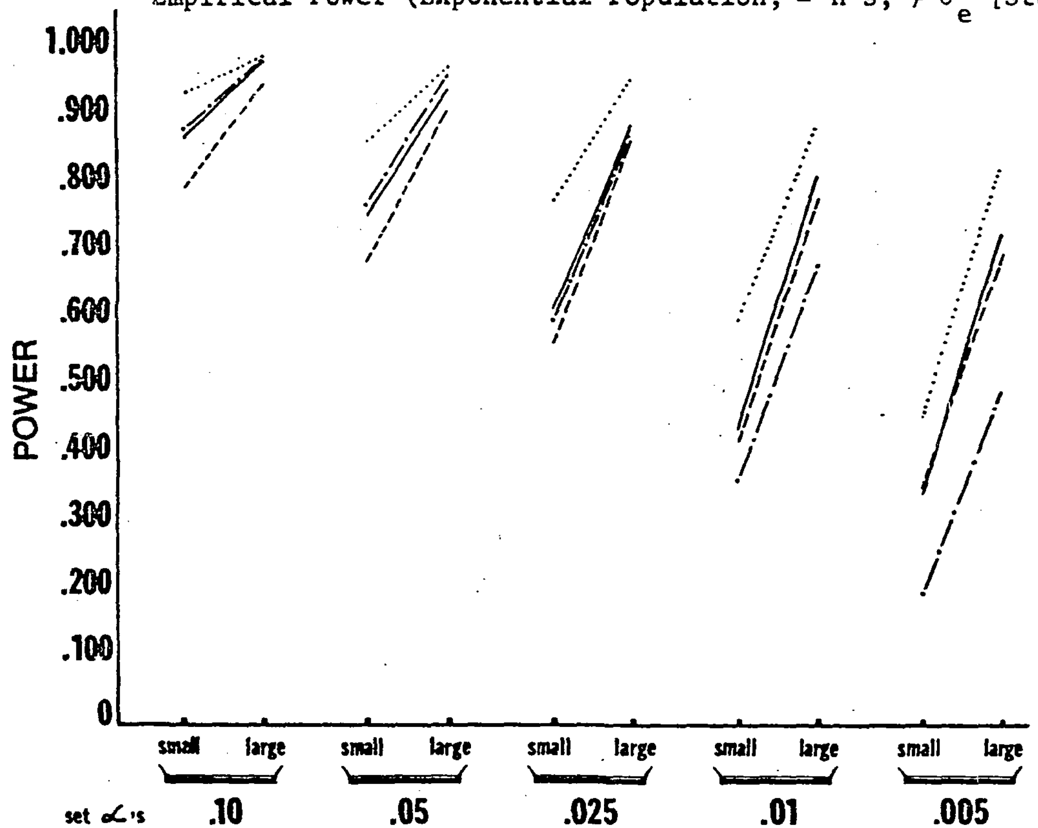
Empirical Power (Normal Population; = n's, $\neq \sigma_e^2$ [stepwise])



LEGEND
———— Power for the ANOVA F-test based on stepwise mean differences
- - - - - Power for the ANOVA F-test based on nonstepwise mean differences
.... Power for the Kruskal-Wallis test based on stepwise mean differences
—·—·— Power for the Kruskal-Wallis test based on nonstepwise mean differences
small Small mean differences (power = .60 for a normal population at .05 level)
large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 2

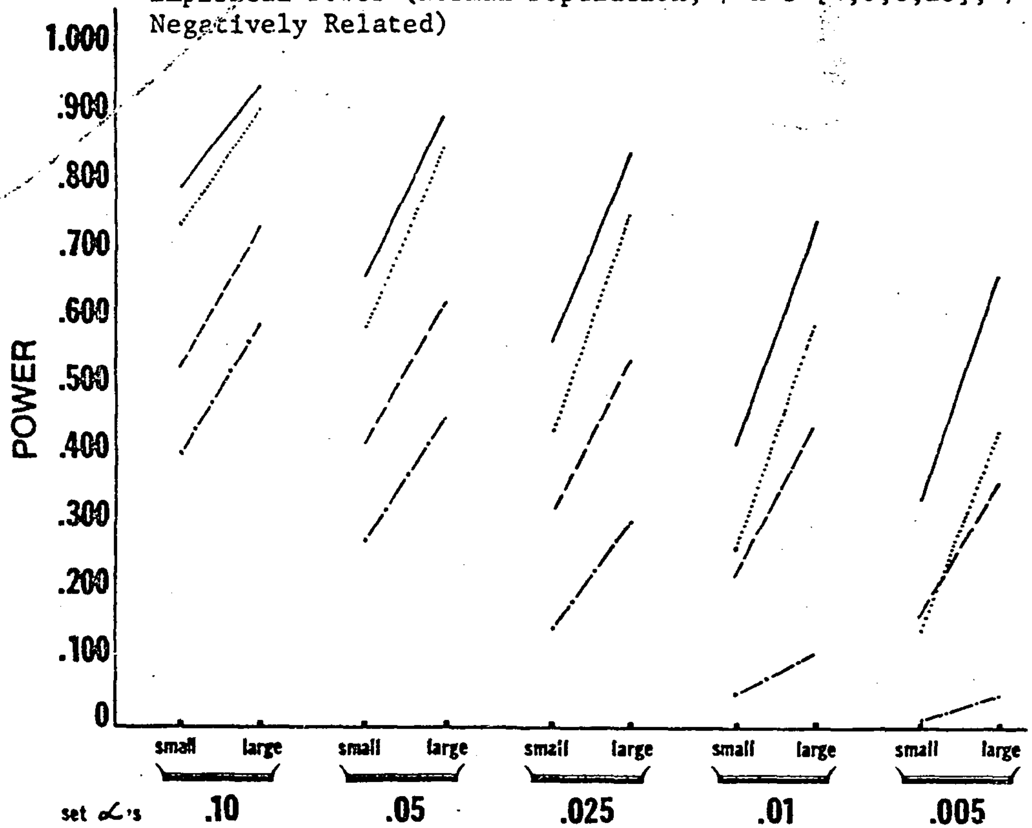
Empirical Power (Exponential Population; = n's, $\neq \sigma_e^2$ [stepwise])



LEGEND
- ——— Power for the ANOVA F-test based on stepwise mean differences
- - - - Power for the ANOVA F-test based on nonstepwise mean differences
- ········ Power for the Kruskal-Wallis test based on stepwise mean differences
- —·—·— Power for the Kruskal-Wallis test based on nonstepwise mean differences
- small: Small mean differences (power = .60 for a normal population at .05 level)
- large: Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 3

Empirical Power (Normal Population; $\neq$ n's [4,6,8,10], $\neq \sigma_e^2$ [stepwise]; Negatively Related)

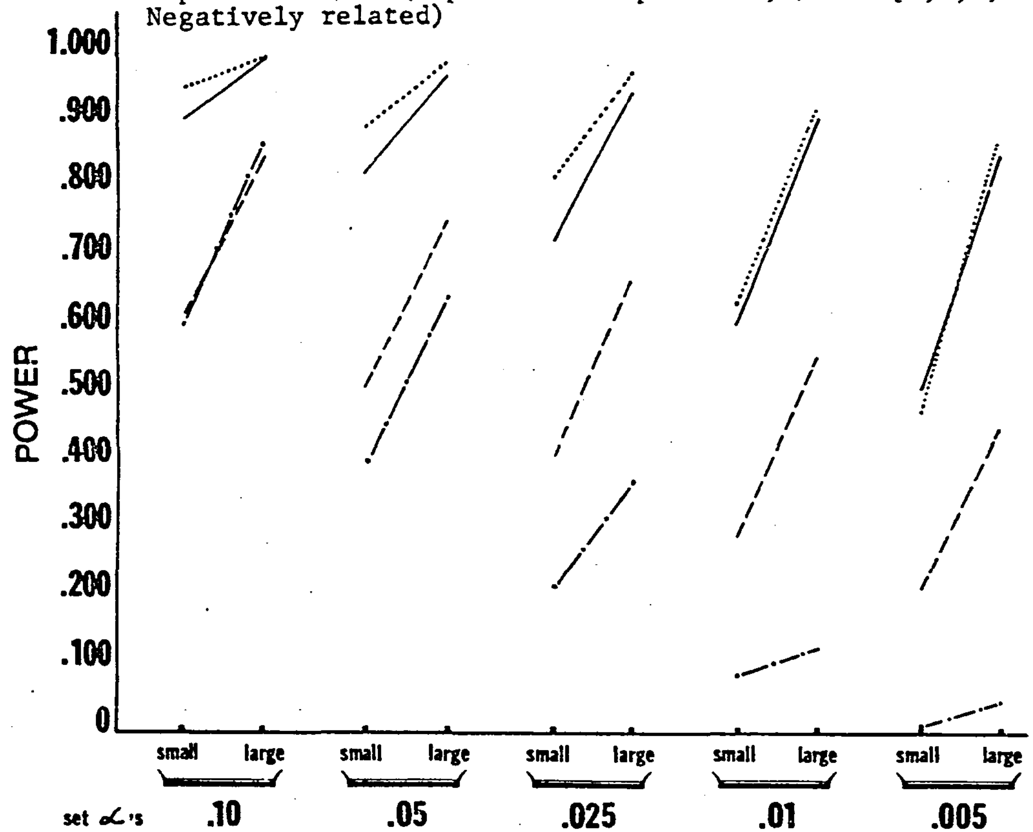


LEGEND

———	Power for the ANOVA F-test based on stepwise mean differences
– – – –	Power for the ANOVA F-test based on nonstepwise mean differences
··········	Power for the Kruskal-Wallis test based on stepwise mean differences
—·—	Power for the Kruskal-Wallis test based on nonstepwise mean differences
small	Small mean differences (power = .60 for a normal population at .05 level)
large	Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 4

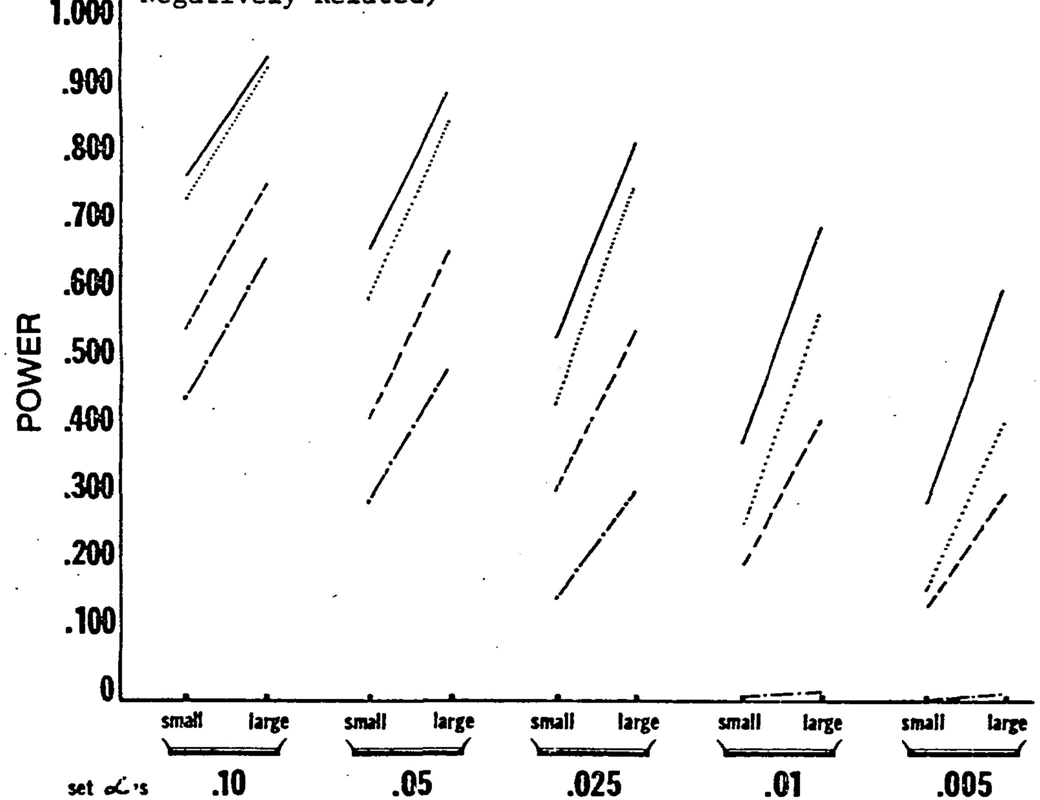
Empirical Power (Exponential Population; ≠ n's [4,6,8,10], ≠ σ_e^2 [stepwise]; Negatively related)



LEGEND
- ——— Power for the ANOVA F-test based on stepwise mean differences
- - - - - Power for the ANOVA F-test based on nonstepwise mean differences
- .... Power for the Kruskal-Wallis test based on stepwise mean differences
- •—•— Power for the Kruskal-Wallis test based on nonstepwise mean differences
- small Small mean differences (power = .60 for a normal population at .05 level)
- large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 5

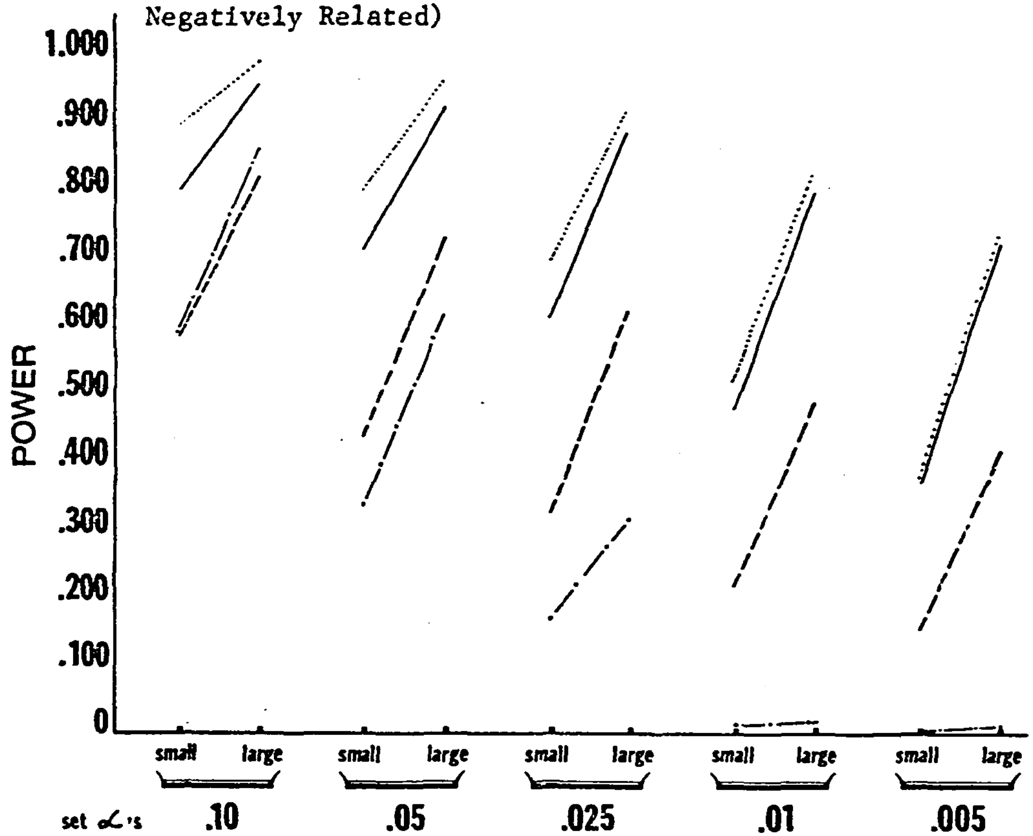
Empirical Power (Normal Population; ≠ n's [4,8,8,8], ≠ σ_e^2 [stepwise]; Negatively Related)



LEGEND
- ——— Power for the ANOVA F-test based on stepwise mean differences
- - - - - Power for the ANOVA F-test based on nonstepwise mean differences
- .... Power for the Kruskal-Wallis test based on stepwise mean differences
- —·— Power for the Kruskal-Wallis test based on nonstepwise mean differences
- small Small mean differences (power - .60 for a normal population at .05 level)
- large Large mean differences (power = .86 for a normal population at .05 level)

FIGURE 6

Empirical Power (Exponential Population; $\neq$ n's [4,8,8,8], $\neq \sigma_e^2$ [stepwise]; Negatively Related)



LEGEND
- ——— Power for the ANOVA F-test based on stepwise mean differences
- – – – Power for the ANOVA F-test based on nonstepwise mean differences
- .... Power for the Kruskal-Wallis test based on stepwise mean differences
- —·— Power for the Kruskal-Wallis test based on nonstepwise mean differences
- small: Small mean differences (power = .60 for a normal population at .05 level)
- large: Large mean differences (power = .86 for a normal population at .05 level)